

RETRIEVING A CONTEXT TREE FROM EEG DATA

BY A. DUARTE^{†,||,‡‡}, R. FRAIMAN^{**}, A. GALVES^{‡,||}, G. OST^{§,||,‡‡} AND C. VARGAS^{¶,‡‡}

*Universidade São Paulo^{||}, Universidad de la República Uruguay^{**},
Universidade Federal do Rio de Janeiro^{††} and Université de
Cergy-Pontoise^{‡‡}*

It has been repeatedly conjectured that the brain retrieves statistical regularities from stimuli, so that their structural features are separated from noise. Here we present a new statistical approach allowing to address this conjecture. This approach is based on a new class of stochastic processes driven by context tree models. Also, it associates to a new experimental protocol in which structured auditory sequences are presented to volunteers while electroencephalographic signals are recorded from their scalp. A statistical model selection procedure for functional data is presented to analyze the electrophysiological signals. This procedure is proved to be consistent. Applied to samples of electrophysiological trajectories collected during structured auditory stimuli presentation, it produces results supporting the conjecture that the brain effectively identifies the context tree characterizing the source.

1. Introduction. This paper can be summarized as follows.

- We present a new statistical approach allowing to address the conjecture that the brain operates as statistician, assigning models to external stimuli.
- This is done through the introduction of a new class of stochastic processes driven by context tree models.

*Updated on January 30

[†]Supported successively by CNPq PhD fellowship grant 141270/2013-6, Post-doctoral FAPESP fellowship grant 2015/11579-5 and CNPq Post-doctoral fellowship grant 201696/2015-0.

[‡]Partially supported by CNPq fellowship grant 309501/2011-3.

[§]Supported successively by CNPq PhD fellowship grant 141482/2013-3, post-doctoral FAPESP fellowship grant 2015/11578-9 and CNPq post-doctoral fellowship grant 201572/2015-0.

[¶]Partially supported by CNPq fellowship grant 306817/2014-4.

MSC 2010 subject classifications: 91E30, (62G20,60K99,62M05)

Keywords and phrases: stochastic processes driven by context tree models, statistical model selection for functional data, stochastic modeling of EEG data, projective method

- This approach leads to a new experimental protocol to test this conjecture.
- In order to analyze the electrophysiological signals collected using our experimental protocol we introduce a new statistical model selection procedure for functional data.
- We also prove that this procedure is consistent. This is exemplified with an application to simulated data reproducing some of the features of the experimental data.
- Applied to samples of electrophysiological trajectories collected by our team this new functional data statistical selection procedure finds results which support the conjecture that the brain effectively identifies the context tree characterizing the source.

Since the pioneer work of [von Helmholtz \(1867\)](#) it has been widely conjectured that the brain does statistical model selection by assigning probabilistic models to samples of stimuli. Contemporary neuroscience has exhaustively attempted to address this fundamental issue (see for instance [Mumford \(1992\)](#), [Wolpert, Ghahramani and Flanagan \(2001\)](#), [Wolpert and Flanagan \(2001\)](#), [Lee and Mumford \(2003\)](#), [Kersten, Mamassian and Yuille \(2004\)](#), [Friston \(2005\)](#), [Summerfield et al. \(2006\)](#), [Körding and Wolpert \(2006\)](#), [Doya et al. \(2007\)](#), [Friston and Kiebel \(2009\)](#), [Tenenbaum et al. \(2011\)](#), [Orbán and Wolpert \(2011\)](#), [Meyer and Olson \(2011\)](#), [Friston \(2012\)](#) and references herein).

Early evidence that the brain performs probabilistic modeling comes mostly from behavioral experiments (see, among others, [Wolpert, Doya and Kawato \(2003\)](#) for a review). There is also evidence that the brain employs predictive coding through the identification of markers of prediction and deviance detection ([Friston \(2005\)](#), [Garrido et al. \(2009\)](#), [Wacongne et al. \(2011\)](#) and [Wacongne, Changeux and Dehaene \(2012\)](#)). However, to the best of our knowledge, none of these papers succeeded in presenting a new statistical framework in which model selection was rigorously employed to test the conjecture that the brain assigns probabilistic models to samples of stimuli. This is precisely what we do in the present paper.

Our approach is based in the introduction of a new class of stochastic processes driven by context models. Using this new class of stochastic processes it is possible to establish a formal relationship between structured sequences of random stimuli and the stochastic process associated to the brain processing of the stimuli.

This new framework suggests a new experimental protocol which can be summarized as follows. A volunteer is exposed to sequences of auditory stimuli. The sequences are generated step by step by a random source. Elec-

troencephalographic (EEG) signals are recorded during the exposure to the auditory sequence of stimuli. The conjecture is that the volunteers' brain automatically identifies the context tree characterizing the source. If this is the case, a signature of the structure of the source should be encoded in the brain activity. The question is whether this signature can be identified in the EEG data recorded during the experiment. This is the challenge faced here.

Considering the electroencephalogram as the realization of a stochastic process means that we have a problem of statistical model selection with functional data which is in general a difficult and unsolved issue. However the stochastic process driven by a context tree model approach adopted here makes the question treatable. This is done by the introduction of a new model selection procedure for functional data driven by a context tree model. This procedure is proved to be consistent. The effective application of the procedure to EEG data uses the projective method introduced in [Cuesta-Albertos, Fraiman and Ransford \(2006\)](#).

Let us describe briefly the structure of the random source generating the auditory stimuli. This source is defined by an algorithm which sets the probabilities of occurrence of each next unit. The algorithm has two components.

- The first component describes a partition of the set of all possible sequences of past units. This partition is defined using only the final past units. The number of units used to define the partition is variable and changes as a function of the past itself.
- The second component is a family of probability measures in the set of auditory units. This family of measures is indexed by the elements of the partition. In other terms, the probability of occurrence of a next unit depends on the element of the partition to which the past up to that point belongs.

Random sources of this type have been introduced by [Rissanen \(1983\)](#) under the name of *context tree models* (CTM). The reason of this name is the following. Partitions of the past as described above can be represented by a rooted and labeled tree. In this tree, each element of the partition is described as a leaf of the tree. Rissanen called each element of the partition, each *leaf*, a *context*.

Introduced by [Rissanen \(1983\)](#) as a universal system for data compression, the context tree models become known in the statistics community through [Bühlmann and Wyner \(1999\)](#) in which they appear with the name of *variable length Markov chains* (VLMC). Context tree models are *stochastic chains with memory of variable length* which is the name adopted in the review

paper by Galves and Löcherbach (2008) to which we refer the reader for a self-contained presentation of the field.

This article is organized as follows. In Section 2 we introduce the notation and basic definitions, including the notion of *context tree model*, presented in Subsection 2.1 and the notion of *stochastic process driven by a context tree model* presented in Subsection 2.3. The random sources used in our experimental protocol are examples of context tree models. They are presented in Subsection 2.2. Our new procedure for statistical model selection is presented in Section 3. Theorem 3.1 on the consistency of the model selection procedure is also presented in Section 3. The case study with data obtained using our experimental protocol is presented in Section 4. In particular in Subsection 4.3 we state Theorem 4.1 on the consistency of the statistical model selection using the projective method. A simulation study is presented in Section 5. The proofs of Theorems 3.1 and 4.1 are presented in Appendices A.1 and A.2 respectively. In Appendix A.3 is proved a technical lemma needed in Subsection 4.4.

2. Stochastic processes driven by context tree models.

2.1. Context tree models. Let A be a finite alphabet. Given two integers $m, n \in \mathbb{Z}$ with $m \leq n$, the string (u_m, \dots, u_n) of symbols in A is often denoted by u_m^n ; its length is $\ell(u_m^n) = n - m + 1$. The empty string is denoted by \emptyset and its length is $\ell(\emptyset) = 0$. Fixed two strings u and v of elements of A , we denote by uv the string in $A^{\ell(u)+\ell(v)}$ obtained by the concatenation of u and v . By definition $u\emptyset = \emptyset u = u$ for any string $u \in A^{\ell(u)}$. The string u is said to be a *suffix* of v if there exists a string s satisfying $v = su$. This relation will be denoted by $u \preceq v$. When $v \neq u$ we say that u is a *proper suffix* of v and write $u \prec v$. Hereafter, the set of all finite strings of symbols in A is denoted by $A^* := \bigcup_{k=1}^{\infty} A^k$.

DEFINITION 1. A finite subset τ of A^* is a *context tree* if it satisfies the following conditions:

1. *Suffix Property.* For no $w \in \tau$ we have $u \in \tau$ with $u \prec w$.
2. *Irreducibility.* No string belonging to τ can be replaced by a proper suffix without violating the suffix property.

The set τ can be easily identified with the set of leaves of a rooted tree with a finite set of labeled branches. The elements of τ will be always denoted by w .

DEFINITION 2. The height of the context tree τ is defined as

$$\ell(\tau) = \max\{\ell(w) : w \in \tau\}.$$

Given a context tree τ , let $p = \{p(\cdot | w) : w \in \tau\}$ be a family of probability measures on A indexed by the elements of τ .

DEFINITION 3. The pair (τ, p) will be called a *probabilistic context tree* on A . Each element of τ will be called a *context*.

DEFINITION 4. Let τ and τ' be two context trees. We say that τ is *smaller* than τ' and write $\tau \preceq \tau'$, if for every $w' \in \tau'$ there exists $w \in \tau$ such that $w \preceq w'$. If $\tau \neq \tau'$ we write $\tau \prec \tau'$ and say that τ is *strictly smaller* than τ' .

DEFINITION 5. Let (τ, p) be a probabilistic context tree on A . A time-homogeneous irreducible stochastic chain $(X_n)_{n \in \mathbb{Z}}$ taking values in A is called a *context tree model* compatible with (τ, p) if

1. For each integer $n \geq \ell(\tau)$ and any finite string $x_{-n}^{-1} \in A^n$ such that $P(X_{-n}^{-1} = x_{-n}^{-1}) > 0$, it holds that

$$(2.1) \quad P(X_0 = a \mid X_{-n}^{-1} = x_{-n}^{-1}) = p(a \mid c_\tau(x_{-n}^{-1})) \quad \text{for all } a \in A,$$

where $c_\tau(x_{-n}^{-1})$ is the only context in τ which is a suffix of x_{-n}^{-1} .

2. No proper suffix of $c_\tau(x_{-n}^{-1})$ satisfies (2.1).

Context tree models will also be called *stochastic chains with memory of variable length*.

2.2. *The random sources used in our experimental protocol.* We can now formally define the two random sources producing the auditory stimuli in our experimental protocol. Both sources are examples of context tree models generating random sequences of *strong beats*, *weak beats* and *silent units*. We use the symbols 0, 1 and 2 to represent respectively silent units, weak beats and strong beats, so that the alphabet of the two sources is the set $A = \{0, 1, 2\}$.

The sources can be described as follows:

- (i) Consider the deterministic sequence $(s_n)_{n \in \mathbb{Z}}$ either

$$\dots 2 \ 1 \ 1 \ 2 \ 1 \ 1 \ 2 \ 1 \ 1 \ 2 \dots$$

or

... 2 1 0 1 2 1 0 1 2 1 0 1 2 ...

(ii) Fix $\epsilon > 0$ and denote by $(\xi_n)_{n \in \mathbb{Z}}$ a sequence of independent and identically distributed Bernoulli random variables such that

$$(2.2) \quad P(\xi_n = 0) = 1 - P(\xi_n = 1) = \epsilon.$$

The parameter ϵ is the omission probability of the weak beats. For each $n \in \mathbb{Z}$ we define

$$X_n = \begin{cases} s_n, & \text{if } s_n \in \{0, 2\} \\ \xi_n & \text{if } s_n = 1. \end{cases}$$

The stochastic chain derived from the first deterministic sequence with period 3 will be called *Ternary random source*. The stochastic chain derived from the second deterministic sequence with period 4 will be called *Quaternary random source*.

It is an easy exercise to check that the Quaternary and Ternary random sources are context tree models with context trees given respectively by

$$(2.3) \quad \tau_{quat} = \{000, 100, 200, 10, 20, 01, 21, 2\} \text{ and } \tau_{ter} = \{00, 10, 20, 01, 11, 21, 2\}.$$

The graphical representation of these trees is given in Figure 1 and the associated family of transition probabilities are presented in Table 1.

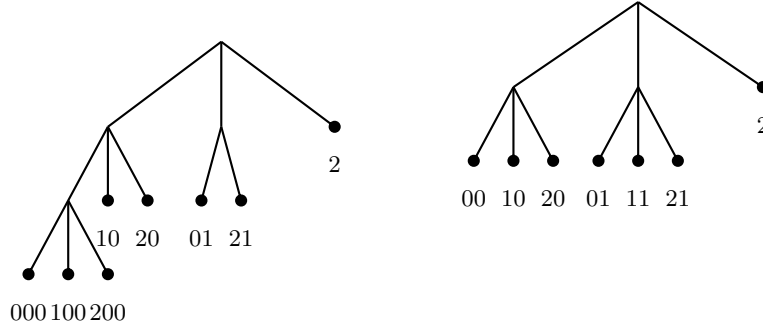


Fig 1: Context trees of the Quaternary (left) and Ternary (right) CTM

2.3. Stochastic processes driven by context tree models. In the experimental protocol described in Section 1, we record a segment of EEG for each single acoustic beat or silent unit generated by the random source. The conjecture is that the distribution of the EEG chunk depends on the context associated to the sequence of acoustic units generated up to the present time step. This suggests the following definition.

Quaternary CTM			Ternary CTM		
context w	p(1 w)	p(2 w)	context w	p(1 w)	p(2 w)
2	$1 - \epsilon$	0	2	$1 - \epsilon$	0
21	0	0	21	$1 - \epsilon$	0
20	0	0	20	$1 - \epsilon$	0
10	$1 - \epsilon$	0	11	0	1
01	0	1	10	0	1
200	$1 - \epsilon$	0	01	0	1
100	0	1	00	0	1
000	0	1			

TABLE 1

Transition probabilities of the Quaternary (left) and Ternary (right) context tree models, for each context.

DEFINITION 6. Let A be a finite alphabet, (τ, p) a probabilistic context tree on A , (F, \mathcal{F}) a measurable space and $(Q^w : w \in \tau)$ a family of probability measures on (F, \mathcal{F}) . The bivariate stochastic chain $(X_n, Y_n)_{n \in \mathbb{Z}}$ taking values in $A \times F$ is a *stochastic process driven by a context tree model* compatible with (τ, p) and $(Q^w : w \in \tau)$ if the following conditions are satisfied,

1. $(X_n)_{n \in \mathbb{Z}}$ is a stochastic chain with memory of variable length compatible with (τ, p) .
2. For any integers $m \leq n$, any string $x_{m-\ell(\tau)+1}^n \in A^{n-m-\ell(\tau)}$ and any sequence J_m, \dots, J_n of \mathcal{F} -measurable sets,

$$P(Y_m \in J_m, \dots, Y_n \in J_n | X_{m-\ell(\tau)+1}^n = x_{m-\ell(\tau)+1}^n) = \prod_{k=m}^n Q^{c_\tau(x_{k-\ell(\tau)+1}^k)}(J_k),$$

where $c_\tau(x_{k-\ell(\tau)+1}^k)$ is the context in τ assigned to the string of symbols $x_{k-\ell(\tau)+1}^k$.

The process $(X_n)_{n \in \mathbb{Z}}$ will be called the *source chain* and $(Y_n)_{n \in \mathbb{Z}}$ will be called the *response chain*.

EXAMPLE 1. Our experimental protocol can be presented in the framework of stochastic processes driven by context tree models. This is done as follows.

- The source chain $(X_n)_{n \in \mathbb{Z}}$ represents the sequence of units of the auditory stimuli. This chain takes values in the alphabet $\{0, 1, 2\}$ where 1 (respectively 2) indicates that the auditory unit is a weak (respectively strong) beat and 0 indicates a silent unit. The chain $(X_n)_{n \in \mathbb{Z}}$ represents either the sequence of the stimuli produced by the Ternary

or the Quaternary random source which have the context trees defined in (2.3) and graphically represented in Figure 1.

- The response chain $(Y_n)_{n \in \mathbb{Z}}$ is constituted by the successive chunks of EEG data, each chunk corresponding to an auditory unit. This means that each Y_n is a real function $Y_n = (Y_n(t), t \in [0, T])$, where T is the time distance between the onsets of two consecutive auditory stimuli. The sample space F is the space $L^2([0, T])$ and \mathcal{F} is the Borel σ -algebra on F .
- Finally, $(Q^w, w \in \tau)$ is a family of probability measures on $L^2([0, T])$.

The goal of the experimental protocol sketched in Section 1 is to produce empirical evidence supporting the conjecture that the brain does statistical model selection. In our case, this means checking whether it is possible to retrieve the structure of the source producing the auditory stimuli from a statistical analysis of the EEG data. Using the formalism of stochastic processes driven by context tree models described above this question can be formulated as follows. Is it possible to recover the context tree τ defining the random source generating the chain $(X_n)_{n \in \mathbb{Z}}$ from the observable sample Y_1, \dots, Y_n ? The statistical framework to address this question is presented in the next section.

3. Statistical model selection. In this section we address the problem of statistical model selection in the class of the stochastic processes driven by context tree models. More precisely, let (F, \mathcal{F}) be a measurable space, $(\bar{\tau}, \bar{p})$ be a probabilistic context tree on a finite alphabet A and $(\bar{Q}^w : w \in \bar{\tau})$ a family of probability measures on (F, \mathcal{F}) . Finally, let $(X_1, Y_1), \dots, (X_n, Y_n)$, with $X_k \in A$ and $Y_k \in F$ for $1 \leq k \leq n$, be a sample produced by a stochastic process driven by the context tree model compatible with $(\bar{\tau}, \bar{p})$ and $(\bar{Q}^w : w \in \bar{\tau})$. Our task is to present a statistical procedure to select a context tree from the sample $(X_1, Y_1), \dots, (X_n, Y_n)$. Before presenting our statistical selection procedure we need two more definitions.

DEFINITION 7. Let τ be a context tree and s a finite string of symbols in A . We define the *branch* in τ induced by s as the set $B_\tau(s) = \{w \in \tau : w = as \text{ for some } a \in A\}$.

Given a sample (X_1, \dots, X_n) of symbols in A and a finite string $u \in A^*$, we write $N_n(u)$ to denote the number of occurrences of the string u in the

sample (X_1, \dots, X_n) , that is

$$N_n(u) = \sum_{m=\ell(u)}^n 1\{X_{m-\ell(u)+1}^m = u\}.$$

DEFINITION 8. Given integers $n > L \geq 1$, an *admissible context tree* of maximal height L , for the sample (X_1, \dots, X_n) of symbols in A , is any context tree τ such that

1. $w \in \tau$ if and only if $\ell(w) \leq L$ and $N_n(w) \geq 1$.
2. Any string $u \in A^*$ with $N_n(u) \geq 1$ is a suffix of some $w \in \tau$ or has a suffix $w \in \tau$.

For any pair of integers $n > L \geq 1$ and any string $u \in A^*$ with $\ell(u) \leq L$, the set of indexes belonging to $\{L, \dots, n\}$ in which the string u appears in sample (X_1, \dots, X_n) is denoted by $I_n(u)$, that is

$$I_n(u) = \{L \leq m \leq n : X_{m-\ell(u)+1}^m = u\}.$$

Observe that by definition $|I_n(u)| = N_n(u)$. If $I_n(u) = \{m_1, \dots, m_{N_n(u)}\}$, for each $m_k \in I_n(u)$, we set $Y_k^{(u)} = Y_{m_k}$, so that $(Y_1^{(u)}, \dots, Y_{N_n(u)}^{(u)})$ is the *subsample of Y_1^n induced by the string u* .

Finally, let D be a measurable function $D : \bigcup_{m_1=1, m_2=1}^\infty F^{m_1} \times F^{m_2} \rightarrow \mathbb{R}_+$ and $(c_n)_{n \geq 1}$ be a sequence of positive real numbers satisfying, for any pair of strings $u, v \in A^*$ with $\max\{\ell(u), \ell(v)\} \leq L$, the following conditions

C.1 if there exists $w \in \bar{\tau}$ such that $w \preceq u$ and $w \preceq v$, it holds

$$\lim_{n \rightarrow \infty} P(D((Y_1^{(u)}, \dots, Y_{N_n(u)}^{(u)}), (Y_1^{(v)}, \dots, Y_{N_n(v)}^{(v)})) > c_n) = 0,$$

C.2 if $u, v \in B_{\bar{\tau}}(s)$ for some $s \in A^*$ and $\bar{Q}_u \neq \bar{Q}_v$, then

$$\lim_{n \rightarrow \infty} P(D((Y_1^{(u)}, \dots, Y_{N_n(u)}^{(u)}), (Y_1^{(v)}, \dots, Y_{N_n(v)}^{(v)})) \leq c_n) = 0.$$

Our selection procedure can be now described as follows. For given integers $1 \leq L < n$, let \mathcal{T}_n be the largest admissible context tree of maximal height L for the sample (X_1, \dots, X_n) . The largest means that if τ is any other admissible context tree of maximal height L for the sample X_1^n , then $\tau \preceq \mathcal{T}_n$. For any string $s \in A^*$ with maximal length and such that $|\mathcal{B}_{\mathcal{T}_n}(s)| \geq 2$, we test the null hypothesis

$$H_0^{(s)} : \mathcal{L}(Y_1^{(as)}, \dots, Y_{N_n(as)}^{(as)}) = \mathcal{L}(Y_1^{(bs)}, \dots, Y_{N_n(bs)}^{(bs)}), \text{ for all } a, b \in A,$$

using the test statistic

$$(3.1) \quad \Delta_n(s) = \max_{a,b \in A} D((Y_1^{(as)}, \dots, Y_{N_n(as)}^{(as)}), (Y_1^{(bs)}, \dots, Y_{N_n(bs)}^{(bs)})).$$

We reject the null hypothesis $H_0^{(s)}$ when $\Delta_n(s) > c_n$. When the null hypothesis $H_0^{(s)}$ is not rejected, we prune the branch $B_{\mathcal{T}_n}(s)$ in \mathcal{T}_n and set as a new candidate tree

$$\mathcal{T}_n = (\mathcal{T}_n \setminus B_{\mathcal{T}_n}(s)) \cup \{s\}.$$

On the other hand, if the null hypothesis $H_0^{(s)}$ is rejected, we keep $B_{\mathcal{T}_n}(s)$ in \mathcal{T}_n and $H_0^{(u)}$ is not tested any longer for all other strings $u \in A^*$ such that $u \preceq s$.

In each pruning step take always the largest string $s \in A^*$ which has not been tested yet. This pruning procedure is repeated until no more pruning is performed. We denote by $\hat{\tau}_n$ the final context tree obtained by this pruning procedure. The formal description of the above procedure is provided in Algorithm 1 as a pseudo code.

Algorithm 1

Input: A sample $(X_1, Y_1), \dots, (X_n, Y_n)$ with $X_n \in A$ and $Y_n \in F$, a measurable function D and the sequence $(c_n)_{n \geq 1}$ fulfilling conditions [C.1](#) and [C.2](#), and a positive integer L .

Output: A tree $\hat{\tau}_n$

```

1:  $\tau \leftarrow \mathcal{T}_n$ 
2:  $\text{Flag}(s) \leftarrow$  “not visited” for all internal node  $s$  of  $\tau$  such that  $|B_\tau(s)| \geq 2$ 
3: for  $k$  in  $L$  to 1 do
4:   while  $\exists s: \ell(s) = k, \text{Flag}(s) =$  “not visited” and  $|B_\tau(s)| \geq 2$  do
5:     take  $s$  such that  $\ell(s) = k, \text{Flag}(s) =$  “not visited” with  $|B_\tau(s)| \geq 2$ 
6:     compute the test statistic  $\Delta_n(s)$  to test  $H_0^{(s)}$ 
7:     if  $\Delta_n(s) > c_n$  then
8:        $\text{Flag}(u) \leftarrow$  “visited”  $\forall u \preceq s$  such that  $|B_\tau(u)| \geq 2$ 
9:     else
10:       $\text{Flag}(s) \leftarrow$  “visited”
11:       $\tau \leftarrow (\tau \setminus B_\tau(s)) \cup \{s\}$ 
12:    end if
13:  end while
14: end for
15: Return  $\hat{\tau}_n = \tau$ .
```

The consistency of Algorithm 1 is the content of our first theorem.

THEOREM 3.1. *Let (X_1^n, Y_1^n) be a sample produced by a stochastic process driven by a context tree model compatible with $(\bar{\tau}, \bar{p})$ and $(\bar{Q}_w : w \in \bar{\tau})$ and let $\hat{\tau}_n$ be the context tree selected from the sample by Algorithm 1 with $L \geq \ell(\tau)$.*

If the statistic D and the threshold sequence $(c_n)_{n \geq 1}$ satisfy conditions [C.1](#) and [C.2](#) then

$$\lim_{n \rightarrow \infty} P(\hat{\tau}_n \neq \bar{\tau}) = 0.$$

4. Case study: retrieving context trees from EEG data. A total of 20 healthy volunteers (9 female, mean age 30, standard deviation 6.8, 18 right handed) was evaluated. All the volunteers did not have any neurological pathology. The volunteers signed a informed consent term, after the nature of the study and the protocol to be performed had been completely understood. This experimental protocol was approved by the local ethical research committee (process number 22047613.2.0000.5261).

4.1. Experimental protocol. The experimental protocol is the one briefly sketched in [Section 1](#). The experiment consisted in exposing volunteers to sequences of auditory stimuli defined as strong beats, weak beats and silent units, indicated respectively by symbols 2, 1 and 0. The sequence of auditory stimuli were produced by the context tree models called in [Subsection 2.2](#) Ternary and Quaternary random sources. Herein, the omission probability ϵ (see [equation \(2.2\)](#)) took the value 0.2.

A third random source was employed to separate the samples generated by the Ternary and Quaternary random sources. This third random source produced sequences of independent auditory stimuli taking the values 0, 1 and 2 with probability 1/3. The corresponding context tree reduces to the single root with no branches. The goal of introducing this unpredictable sequence was to *shuffle cards* before the volunteer is exposed to a next sample.

The volunteer was exposed to two 12 min blocks of samples generated by each of the random sources. The blocks were separated from each other by a period of time ranging from 5 to 10 min, during which data collection was interrupted. Each sample was a concatenation of three 1 min sequences of auditory units generated independently by the same random source. Each sequence of auditory units was separated from the next one by a 15 seconds silent interval.

All volunteers were exposed to two different orderings, either Ternary, Independent, Quaternary or Quaternary, Independent, Ternary. For half of the volunteers the starting block was Ternary, Independent, Quaternary and the second block was Quaternary, Independent and Ternary. The inverse ordering was used with the other half, to balance possible order effects.

Presentation software (Presentation Mixer as a Primary Buffer and a Sound card: SoundMAX HD Audio) was used to play the auditory sequences through a headset. The loudness of the stimuli was individually regulated

before the experiment start by presenting the strong beat and asking the volunteer to adjust it up to a comfortable level (Range: 0.1-0.3 dB).

4.2. Data acquisition and pre-processing. Electroencephalographic recording was performed by means of a 128 channels system (Geodesic HydroCel GSN 128 EGI, Electrical Geodesic Inc.) during the exposure to the auditory sequence of stimuli. The electrode cap, previously immersed in saline solution (KCl), was dressed into the volunteer's scalp. Volunteers were instructed to close their eyes and remain quiet throughout the experiment.

The EEG signal was amplified with a nominal gain of 20 times. The acquisition was performed in a sampling frequency of 250Hz. During acquisition the signal was analogically filtered (Butterworth first order band-pass filter of 0.1-200Hz; Geodesic EEG System 300, Electrical Geodesic Inc.). The electrode positioned on the vertex (Cz) was used as reference.

The data was preprocessed offline using EEGLAB (Delorme and S.Makeig (2004)) running in MATLAB environment (MathWorks, Natick, MA, version R2012a). Signals were filtered with a Butterworth fourth order band-pass filter of 1-30 Hz. Artifacts above and below 100 μV were removed. The data was then segmented into events of 450ms, each one indexed by the corresponding auditory unit. Finally, baseline correction was performed using the signal collected 50 ms before each event start (see figure 2).

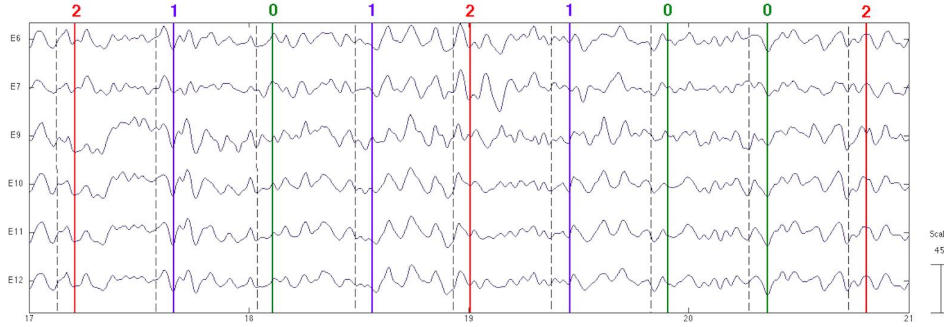


Fig 2: EEG time evolution recorded in seven electrodes during the exposure to auditory units produced by the Quaternary source. Solid vertical lines indicate the beginning of the successive auditory units indexed by the symbols 2, 1 and 0. Dashed lines indicate the boundaries of successive chunks of EEG signals constituting the sequence $(Y_n)_{n \in \mathbb{Z}}$. The distance between each dashed line and the next solid line is 50ms. The amplitude scale of 45 μV is indicated as a vertical interval on right-hand side of the figure.

4.3. *Statistical framework.* In this section we discuss how to apply the theoretical framework presented in Section 3 for the case where the driven process $(Y_n)_{n \in \mathbb{Z}}$ is a sequence of chunks of EEG signals produced by our experimental protocol as described in Example 1.

The framework is the following. The stochastic process driven by a context tree model $(X_n, Y_n)_{n \in \mathbb{Z}}$ is such that $Y_n \in L^2([0, T])$ for some $T > 0$. To use Theorem 3.1 we need to define a statistic for samples of elements of $L^2([0, T])$ and a threshold sequence $(c_n)_{n \geq 1}$ satisfying both conditions C.1 and C.2. This will be done by using the projective method (Cuesta-Albertos, Fraiman and Ransford (2006)) as follows.

Let $1 \leq L < n$ be positive fixed integers and $(X_1, Y_1), \dots, (X_n, Y_n)$ be a sample produced by a stochastic process driven by a context tree model with $Y_k \in L^2([0, T])$, for $1 \leq k \leq n$, for some $T > 0$. Given a string $u \in A^*$ with $\ell(u) \leq L$ and an element $h \in L^2([0, T])$ define the empirical distribution associated to string u and direction h as

$$\hat{Q}_n^{u,h}(t) := \frac{1}{N_n(u)} \sum_{m=1}^{N_n(u)} 1_{(-\infty, t]}(\langle Y_m^{(u)}, h \rangle), \quad t \in \mathbb{R},$$

where for any pair of functions $f, h \in L^2([0, T])$,

$$\langle f, h \rangle = \int_0^T f(t)h(t)dt.$$

For a given pair of finite strings u and v with $\max\{\ell(u), \ell(v)\} \leq L$ and $h \in L^2([0, T])$, the Kolmogorov-Smirnov distance between the empirical distributions $\hat{Q}_n^{u,h}$ and $\hat{Q}_n^{v,h}$ is denoted by

$$KS(\hat{Q}_n^{u,h}, \hat{Q}_n^{v,h}) := \sup_{t \in \mathbb{R}} |\hat{Q}_n^{u,h}(t) - \hat{Q}_n^{v,h}(t)|.$$

Let us then define for any pair of strings $w, s \in A^*$ with $\{\ell(u), \ell(v)\} \leq L$ and $h \in L^2([0, T])$,

$$D^h((Y_1^{(u)}, \dots, Y_{N_n(u)}^{(u)}), (Y_1^{(v)}, \dots, Y_{N_n(v)}^{(v)})) = \sqrt{\frac{N_n(u)N_n(v)}{N_n(u) + N_n(v)}} KS(\hat{Q}_n^{u,h}, \hat{Q}_n^{v,h}).$$

Under some assumptions on the family of probability measures generating the response process $(Y_n)_{n \in \mathbb{Z}}$, it will be proved that if the direction h appearing in D^h is chosen as a realization of a Brownian motion $W = (W(t) : 0 \leq t \leq T)$ then the statistic D^W satisfies conditions (C.1) and (C.2) for almost all realizations $W = (W(t) : 0 \leq t \leq T)$. These are the contents of Propositions 1 and 2 in the Appendix A.2.

For this reason, for any string $s \in A^*$ and any realization of a Brownian motion W , we define

$$(4.1) \quad \Delta_n(s) := \Delta_n^W(s) = \max_{a,b \in A} D^W((Y_1^{(as)}, \dots, Y_{N_n(as)}^{(as)}), (Y_1^{(bs)}, \dots, Y_{N_n(bs)}^{(bs)})).$$

To state the consistency theorem in this framework we need the following assumption.

ASSUMPTION 1. The associated family $(\bar{Q}^w : w \in \bar{\tau})$ of probability measures on $(L^2([0, T]), \mathcal{B}(L^2([0, T])))$ is non-atomic and, for any string $s \in A$ such that $|B_{\bar{\tau}}(s)| \geq 2$, there exist $w, w' \in B_{\bar{\tau}}(s)$ such that $\bar{Q}^w \neq \bar{Q}^{w'}$ and either \bar{Q}^w or $\bar{Q}^{w'}$ satisfy the Carleman condition.

For the sake of completeness we recall the *Carleman condition*, as well as the notion of family of *non-atomic* probability measures for probability measures on $(L^2([0, T]), \mathcal{B}(L^2([0, T])))$. A probability measure P defined on $(L^2([0, T]), \mathcal{B}(L^2([0, T])))$ is said to satisfy the *Carleman condition* if all the absolute moments $m_n := \int \|h\|^n P(dh)$, $n \geq 1$ are finite and

$$\sum_{n \geq 1} m_n^{-1/n} = \infty.$$

A probability measure P on $(L^2([0, T]), \mathcal{B}(L^2([0, T])))$ is *non-atomic* if for any $h \in L^2([0, T])$, the distribution P^h is non-atomic. Let V be a finite set of indexes and $(P_i : i \in V)$ be a family of probability measures on $(L^2([0, T]), \mathcal{B}(L^2([0, T])))$. We say that $(P_i : i \in V)$ is non-atomic if for all $i \in V$, the probability measure P_i is non-atomic.

THEOREM 4.1. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a sample produced by a stochastic process driven by a context tree model compatible with $(\bar{\tau}, \bar{p})$ and $(\bar{Q}_w : w \in \bar{\tau})$ and let $\hat{\tau}_n$ be the context tree selected from the sample by Algorithm 1 with $L \geq \ell(\tau)$. If $(\bar{Q}_w : w \in \bar{\tau})$ satisfies Assumption 1 and the threshold sequence $(c_n)_{n \geq 1}$ is such that $c_n \rightarrow \infty$ and $c_n n^{-1/2} \rightarrow 0$ as $n \rightarrow \infty$, then

$$\lim_{n \rightarrow \infty} P(\hat{\tau}_n \neq \bar{\tau}) = 0.$$

The proof of Theorem 4.1 is postponed to Appendix A.2.

4.4. Data analysis. In what follows the application of the statistical procedure described in Section 4.3 is detailed. The analysed data set may be summarized as follows. Let $\mathcal{V} = \{v_1, v_2, \dots, v_{20}\}$ denote the set of 20 volunteers under consideration, $\mathcal{E} = \{e_1, e_2, \dots, e_{18}\}$ the set of all electrodes belonging to the system 10-20 but the reference electrode C_z , and $\mathcal{S} = \{\text{Ternary},$

Quaternary} the set of possible random sources. In the sequel, to avoid a cumbersome notation, we shall describe the data set as well as the application of the statistical procedure used to select a context tree from this data set for a fixed volunteer $v \in \mathcal{V}$ and random source $s \in \mathcal{S}$.

The sample of the auditory sequence produced by the random source $s \in \mathcal{S}$ exposed to the volunteer $v \in \mathcal{V}$ is then denoted by

$$X_1, \dots, X_n$$

where n is the sample size. Since the time of exposure (which is the same for all volunteers) of each random source is 6 min and the length of the interval between two consecutive auditory units is 450ms, it follows that the sample size $n = 800$. The corresponding EEG recording restricted to the set of electrodes \mathcal{E} is denoted by

$$Y_1, \dots, Y_n,$$

where $[0, T] \ni t \mapsto Y_k(t) = (Y_k^e(t) : e \in \mathcal{E})$, with $T = 450ms$, is the vector of EEG chunks associated to the k -th auditory unit X_k , with $1 \leq k \leq n$. Being $250Hz$ the sampling frequency of the EEG acquisition and $T = 450ms$, all the EEG chunks have 113 values. This requires that all the computations necessary to numerically implement our statistical procedure are done with discrete time $t = 1, \dots, 113$. Therefore, the sample data set is

$$(4.2) \quad (X_1, Y_1) \dots, (X_n, Y_n),$$

where $n = 800$ and $Y_k \in \mathbb{R}^{113 \times 18}$, for all $1 \leq k \leq n$.

An extension of the Algorithm 1 (henceforward called Algorithm 2) is then employed to select a context tree from the sample appearing in (4.2). The algorithm 2 works as follows. For each electrode $e \in \mathcal{E}$, apply Algorithm 1 to the data $(X_1, Y_1^e) \dots, (X_n, Y_n^e)$ replacing the steps 6, 7 and 8 by the following procedure. Hereafter fix an integer $N \geq 1$.

- Generate N independent realizations W_1, \dots, W_N of the Brownian motion $W = (W(t) : t \in [0, T])$.
- Compute the test statistics $\Delta_n^{W_1}(s), \dots, \Delta_n^{W_N}(s)$.
- Define the statistic

$$\tilde{\Delta}_n(s) = \sum_{m=1}^N 1\{\Delta_n^{W_m}(s) > c_n\}$$

and if $\tilde{\Delta}_n(s)$ is large, that is $\tilde{\Delta}_n(s) > C_1$ for some constant $C_1 > 0$, perform step 8 of Algorithm 1. Otherwise, perform steps 10 and 11 of Algorithm 1.

Using more than one random projection we increase the power of the test. We denote by $\hat{\tau}_n^e$ the context tree selected from the sample $(X_1, Y_1^e) \dots, (X_n, Y_n^e)$ by applying this powerful version of Algorithm 1. Finally, setting for any $w \in A^*$,

$$Z_n(w) = \sum_{e \in E} 1\{w \in \hat{\tau}_n^e\},$$

the context tree assigned $\hat{\tau}_n$ to sample (4.2) is then defined as

$$\hat{\tau}_n = \{w \in A^* : Z_n(w) > C_2\},$$

for some constant $C_2 > 0$.

The threshold values c_n , C_1 and C_2 are chosen in such a way

$$P(\bar{\tau} \preceq \hat{\tau}_n) \leq 0.1,$$

where $\bar{\tau}$ is the context tree defining the stochastic chain $(X_n)_{n \in \mathbb{Z}}$. The proof of the existence of such a triple is the content of Lemma 1. Its proof is postponed to Appendix A.3. The boundedness of the EEG signals (after the pre-processing step) implies that their underlying family of probability distributions satisfies Carleman condition. Thus the consistency of Algorithm 2 follows immediately from Theorem 4.1.

4.5. Results. The results obtained from the analysis of the EEG recorded while the volunteers were exposed to the Quaternary and Ternary sources are summarized in Figures 3 and 4.

In the Quaternary case this summary shows that

- 15 of the context trees correctly have 2 as a context;
- 19 of the context trees correctly do not have 1 as a context;
- 18 of the context trees correctly do not have 0 as a context;
- 12 of the context trees correctly satisfy the 3 features above. This means that 12 of the twenty volunteers successfully identify the correct context tree truncated at height 2.

In the Ternary case the summary shows that

- only 3 of the context trees correctly have 2 as a context;
- 16 of the context trees correctly do not have 1 as a context;
- 11 of the context trees correctly do not have 0 as a context.

A main reason for context misidentifications in the data analysis could be the small set of samples of EEG signals used in each *prune-or-keep* decision. For instance in the case of the Quaternary source, 14 volunteers failed to identify 000 as a context and 7 volunteers incorrectly decide not to prune the

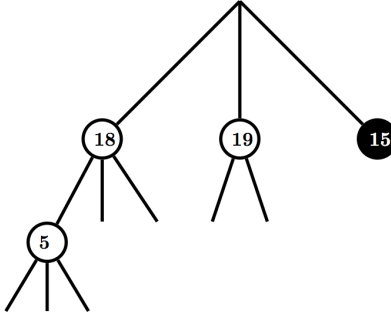


Fig 3: **Context tree summary for the Quaternary data.** This context tree summarizes the twenty context trees obtained from the EEG data recorded while the volunteers are exposed to the Quaternary source. White nodes indicate the number of subjects which correctly identify the node as not being a context. Black nodes indicate the number of subjects which correctly identify the node as a context.

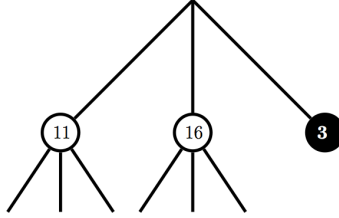


Fig 4: **Context tree summary for the Ternary data.** This context tree summarizes the twenty context trees obtained from the EEG data recorded while the volunteers are exposed to the Ternary source. White nodes indicate the number of subjects which correctly identify the node as not being a context. Black nodes indicate the number of subjects which correctly identify the node as a context.

branch 020. This is not surprising when we observe that for each volunteer, the strings 000 and 020 appear at most 14 and 15 times respectively in the sample. In the Ternary source, the strings 020 and 002 appear at most 17 and 15 times with average number occurrences equal to 10.3 and 10.7 respectively per volunteer.

In general, the results obtained in the Ternary case are less impressive than those obtained for the Quaternary case. This could be a consequence of the fact that in the Ternary case the two random units (which could take the values 0 or 1) occur successively, while in the Quaternary case the two

random units are separated by deterministic units.

In general, the question of the importance of the specific features of the context trees used in our experimental setup should be better discussed. This will be done in forthcoming article.

5. Simulation study. We perform simulation studies for stochastic processes driven by Quaternary and Ternary context tree models presented in Figure 1 and Table 1. Throughout this section, $\bar{\tau}$ is either the Quaternary or Ternary context tree. For each of the two context tree models we consider two different possibilities, P1 and P2 to define the family of probability measures $(\bar{Q}^w : w \in \bar{\tau})$.

If we chose a random function $h \in L^2([0, T])$ according to a probability measure Q on $L^2([0, T])$ and then project h on a realization of a Brownian motion $W = (W(t) : t \in [0, T])$, the resulting variable has a normal distribution with mean 0 and whose variance is given by

$$(5.1) \quad V(h) = \int_0^T \int_0^T \min\{s, t\} h(s)h(t)dsdt.$$

In particular, the elements of set $\mathcal{V}_w = \{V_1, \dots, V_{N_n(w)}\}$, where for each $1 \leq k \leq N_n(w)$, $V_k = V(Y_k^{(w)})$ (recall the definition of $Y_k^{(w)}$ in Subsection 3) have normal distribution whose variance depends on the context w through the distribution $Q^w \circ V^{-1}$. Although the distribution Q^w is unknown we always can compute the set of variances through formula (5.1) from our data set. This set of variances is used to define the two possibilities $P1$ and $P2$ considered in the simulation study. In this way, we incorporate some of the features of the experimental data in the simulated data.

For each context tree model, we choose an electrode of a specific volunteer. For either model the electrode and volunteer chosen were the ones with best performance in the statistical analysis. In the simulation associated to the Quaternary context tree model EEG signals recorded from the electrode $E58$ of the volunteer $V19$ were used. In the Ternary case, the electrode $E92$ of volunteer $V05$ was chosen. For each context $w \in \bar{\tau}$ we compute, using (5.1), the set of variances \mathcal{V}_w of the EEG signals associated to w .

The choices $P1$ and $P2$ defining the family of probability measures $(\bar{Q}^w : w \in \bar{\tau})$ are given now. In $P1$, for each context $w \in \bar{\tau}$, \bar{Q}^w is a Gaussian distribution where the mean μ_w is the average value over the set \mathcal{V}_w , and the variance is $\sigma^2 = 10^{-4}$. The right (left) side of the Table 2 shows the empirical averages of \mathcal{V}_w , for each context $w \in \bar{\tau}$, in the Quaternary (Ternary) case. In $P2$, for each $w \in \bar{\tau}$, we take Q^w as the Uniform distribution over the set \mathcal{V}_w . All the algorithms are provided as supplementary materials.

Transition probabilities of the response chain on P1			
V05 Electrode 92		V19 Electrode 58	
context(w)	μ_w	context(w)	μ_w
2	1.2721	2	3.4438
21	1.2699	21	3.1380
11	1.4476	01	3.5653
01	0.8367	20	3.6105
20	1.0585	10	3.7674
10	1.2667	200	4.7640
00	1.2264	100	2.9685
		000	3.5353

TABLE 2

Transition probabilities of the Quaternary (right) and Ternary (left) stochastic processes in the P1 protocol, when $Q^w = \mathcal{N}(\mu_w, 10^4)$ and μ_w is the empirical average of the set of variances of the EEG signals associated to each context

We considered four choices of samples size n : 50 000, 100 000, 150 000 and 200 000 for each one of the two scenarios, P1 and P2. For each value of n we simulated 100 samples. Then, for each sample we selected a context tree according the statistical model selection described in Section 3, starting from the complete admissible context tree of maximal length 3.

The right (left) side of the Tables 3 and 4, indicates the number of times that, within the 100 simulated samples, the Quaternary (Ternary) context tree was correctly selected by the algorithm under the possibilities P1 and P2 respectively. We also report the number of times the second most selected context tree was chosen by the algorithm. In the Quaternary case in all simulation protocols the second most selected context tree is the Quaternary context tree with the contexts $\{21, 01\}$ replaced by the symbol 1. In the Ternary case, in all simulation protocols the second most selected context tree is the one which has the symbol 0 as a context, instead of $\{00, 10, 20\}$, and keeps all the other contexts of the Ternary context tree.

APPENDIX A: MATHEMATICAL PROOFS

A.1. Proof of Theorem 3.1.

PROOF OF THEOREM 1. Define $C_{\bar{\tau}} = \{w \in \bar{\tau} : |B_{\bar{\tau}}(\text{suf}(w))| \geq 2\}$ where $\text{suf}(w) \in A^*$ is the string obtained from the string w by removing its last symbol. If $w = w_{-k}^{-1}$ for some $k \geq 1$, then $\text{suf}(w) = w_{-(k-1)}^{-1}$ with the convention that $w_0^{-1} = \emptyset$. Define also the following events

$$U_n = \bigcup_{w \in C_{\bar{\tau}}} \{\Delta_n(\text{suf}(w)) \leq c_n\} \text{ and } O_n = \bigcup_{w \in \bar{\tau}} \bigcup_{\substack{s \succ w: \\ \ell(s) \leq L}} \{\Delta_n(s) > c_n\}.$$

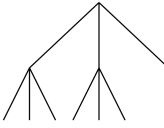
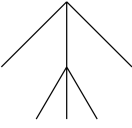
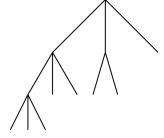
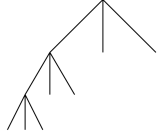
Number of times each context tree was selected in 100 repetitions in P1					
Ternary proportions			Quaternary proportions		
Sample size			Sample size		
50.000	11	82	50.000	30	63
100.000	43	54	100.000	84	12
150.000	74	15	150.000	96	1
200.000	90	3	200.000	96	0

TABLE 3

Number of times, within 100 repetitions, in which each context tree was selected by the algorithm in the Ternary (left) and Quaternary (right) hidden processes under the P1 protocol

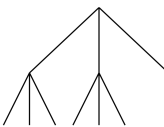
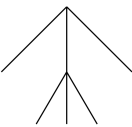
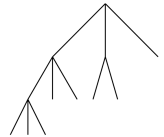
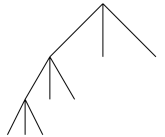
Number of times each context tree was selected in 100 repetitions in P2					
Ternary proportions			Quaternary proportions		
Sample size			Sample size		
50.000	7	88	50.000	26	36
100.000	31	64	100.000	84	9
150.000	71	25	150.000	95	0
200.000	97	0	200.000	99	0

TABLE 4

Number of times, within 100 repetitions, in which each context tree was selected by the algorithm in the Ternary (left) and Quaternary (right) hidden processes under the P2 protocol

It follows from the definition of the Algorithm 1 that $P(\hat{\tau}_n \neq \bar{\tau}) = P(U_n) + P(O_n)$. Thus, we need to prove that both $P(U_n)$ and $P(O_n)$ converge to 0 as $n \rightarrow \infty$. We start by proving that $P(U_n) \rightarrow 0$ as $n \rightarrow \infty$. By the union bound, we see that

$$(A.1) \quad P(U_n) \leq \sum_{w \in \bar{\tau}} P(\Delta_n(\text{suf}(w)) \leq c_n).$$

Since $|B_{\bar{\tau}}(\text{suf}(w))| \geq 2$, there exists a pair $u, v \in B_{\bar{\tau}}(\text{suf}(w))$ whose the associated distributions \bar{Q}^u and \bar{Q}^v on F are different. Observing that

$$\{\Delta_n(\text{suf}(w)) \leq c_n\} \subset \{D((Y_1^{(u)}, \dots, Y_{N_n(u)}^{(u)}), (Y_1^{(v)}, \dots, Y_{N_n(v)}^{(v)}) \leq c_n\},$$

the Condition C.2 then implies that $P(U_n) \rightarrow 0$ as $n \rightarrow \infty$.

To conclude the proof it remains to show that $P(O_n) \rightarrow 0$ as $n \rightarrow \infty$. Using again the union bound, we have

$$(A.2) \quad P(O_n) \leq \sum_{w \in \bar{\tau}} \sum_{\substack{s \succ w: \\ \ell(s) \leq L}} P(\Delta_n(s) > c_n).$$

Observing that

$$\{\Delta_n(s) > c_n\} = \bigcup_{a, b \in A} \{D((Y_1^{(as)}, \dots, Y_{N_n(as)}^{(as)}), (Y_1^{(bs)}, \dots, Y_{N_n(bs)}^{(bs)})) > c_n\},$$

we deduce from Condition [C.1](#) and inequality [\(A.2\)](#) that $P(O_n) \rightarrow 0$ as $n \rightarrow \infty$. \square

A.2. Proof of Theorem 4.1. Theorem 4.1 will follow easily from the two propositions below. The first proposition states that if the family $\{\bar{Q}^w : w \in \bar{\tau}\}$ is continuous and the threshold sequence $(c_n)_{n \geq 1}$ satisfies $c_n \rightarrow \infty$ as $n \rightarrow \infty$, we have that for each $h \in L^2([0, T])$, the statistic D^h and $(c_n)_{n \geq 1}$ fulfill the condition [C.1](#). More precisely,

PROPOSITION 1. *Let $\bar{\tau}$ be a context tree and $\{\bar{Q}^w : w \in \bar{\tau}\}$ be a family of continuous probability distributions on (F, \mathcal{F}) . For any positive integer $L \geq \ell(\tau)$, context $w \in \bar{\tau}$, strings $u, v \in \bigcup_{k=1}^{L-\ell(w)} A^k$ such that $w \preceq u$ and $w \preceq v$, $h \in F \setminus \{0\}$ and $(c_n)_{n \geq 1}$ such that $c_n \rightarrow \infty$ as $n \rightarrow \infty$, it holds that*

$$\lim_{n \rightarrow \infty} P(D^h((Y_1^{(u)}, \dots, Y_{N_n(u)}^{(u)}), (Y_1^{(v)}, \dots, Y_{N_n(v)}^{(v)})) > c_n) = 0.$$

PROOF. Indeed, by recurrence the random variables $N_n(u)$ and $N_n(v)$ diverge as $n \rightarrow \infty$. Hence, Theorem 3.1(a) of [Cuesta-Albertos, Fraiman and Ransford \(2006\)](#) implies that the distribution of the random variable $D^h((Y_1^{(u)}, \dots, Y_{N_n(u)}^{(u)}), (Y_1^{(v)}, \dots, Y_{N_n(v)}^{(v)}))$ is independent of the strings u and v , and also of the direction $h \in F \setminus \{0\}$. Moreover, for each $t > 0$, it holds that

$$(A.3) \quad \lim_{n \rightarrow \infty} P(D^h((Y_1^{(u)}, \dots, Y_{N_n(u)}^{(u)}), (Y_1^{(v)}, \dots, Y_{N_n(v)}^{(v)})) > t) := G(t),$$

where $G(t) = 2 \sum_{k=1}^{\infty} (-1)^{k+1} e^{-2k^2 t^2}$. Given $\epsilon > 0$, let t_0 be a positive real number such that $G(t) < \epsilon/2$ for all $t \geq t_0$. Such t_0 exists since $t \mapsto G(t)$ is an non-decreasing function such that $G(t) \rightarrow 0$ as $t \rightarrow \infty$. Hence, for all n sufficiently large so that $c_n \geq t_0$ and

$$|P(D^h((Y_1^{(u)}, \dots, Y_{N_n(u)}^{(u)}), (Y_1^{(v)}, \dots, Y_{N_n(v)}^{(v)})) > t_0) - G(t_0)| < \epsilon/2.$$

As a consequence for all n large enough

$$P(D^h((Y_1^{(u)}, \dots, Y_{N_n(u)}^{(u)}), (Y_1^{(v)}, \dots, Y_{N_n(v)}^{(v)})) > c_n) < \epsilon,$$

which concludes the proof. \square

We will now show that under Assumption 1 and assuming that $(c_n)_{n \geq 1}$ is such that $c_n \rightarrow \infty$ and $c_n n^{-1/2} \rightarrow 0$ as $n \rightarrow \infty$, for almost all realizations of a Brownian motion $W = (W(t) : t \in [0, T])$, the statistic D^W and the threshold sequence $(c_n)_{n \geq 1}$ also fulfill Condition C.2. This is the content of the second proposition.

PROPOSITION 2. *For any $(c_n)_{n \geq 1}$ satisfying $c_n \rightarrow \infty$ and $c_n n^{-1/2} \rightarrow 0$ as $n \rightarrow \infty$, and any family $\{\bar{Q}^w : w \in \bar{\tau}\}$ of probability distributions on $(L^2([0, T]), \mathcal{B}(L^2([0, T])))$ verifying Assumption 1, there exists a pair $w, w' \in \bar{\tau}$ such that for almost all realization of a Brownian motion $W = (W(t) : t \in [0, T])$ on $[0, T]$,*

$$\lim_{n \rightarrow \infty} P(D^W((Y_1^{(w)}, \dots, Y_{N_n(w)}^{(w)}), (Y_1^{(w')}, \dots, Y_{N_n(w')}^{(w')})) \leq c_n) = 0.$$

PROOF. For each $n \geq 1$ and $h \in L^2([0, T]) \setminus \{0\}$, define

$$N_n := \sqrt{\frac{N_n(w) + N_n(w')}{N_n(w)N_n(w')}} \text{ and } Z_n^h := \frac{\delta_n}{KS(\hat{Q}_n^{h,w}, \hat{Q}_n^{h,w'})} N_n,$$

if $\min\{N_n(w), N_n(w')\} \geq 1$ and $KS(\hat{Q}_n^{h,w}, \hat{Q}_n^{h,w'}) > 0$. Otherwise, we set $N_n = Z_n^h = 0$. Notice that for any $h \in L^2([0, T]) \setminus \{0\}$,

$$\{D^W((Y_1^{(w)}, \dots, Y_{N_n(w)}^{(w)}), (Y_1^{(w')}, \dots, Y_{N_n(w')}^{(w')})) \leq c_n\} = \{Z_n^h \geq 1\},$$

so that the result will follow immediately once we prove that for almost all realization of a Brownian motion W on $[0, T]$, P-almost surely $Z_n^W \rightarrow 0$ as $n \rightarrow \infty$.

By the strong law of large number for context tree models, we have P-almost surely

$$(A.4) \quad \lim_{n \rightarrow \infty} n^{1/2} N_n = \frac{p(w) + p(w')}{p(w)p(w')}.$$

Now Theorem 3.1(b) of Cuesta-Albertos, Fraiman and Ransford (2006) implies that for almost all realization of a Brownian motion W on $L^2([0, T])$, P-almost surely

$$(A.5) \quad \liminf_{n \rightarrow \infty} KS(\hat{Q}_n^{h,w}, \hat{Q}_n^{h,w'}) > 0.$$

Since $c_n n^{-1/2} \rightarrow 0$ as $n \rightarrow \infty$, it follows from (A.4) and (A.5) that for almost all realization of a Brownian motion W on $[0, T]$, P-almost surely

$$Z_n^B = \frac{(n^{-1/2} \delta_n)}{KS(\hat{Q}_n^{B,w}, \hat{Q}_n^{B,w'})} (n^{1/2} N_n) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

□

The proof of Theorem 4.1 is now easy.

PROOF OF THEOREM 4.1.. Indeed, proceed as in the proof of Theorem 3.1. Propositions 1 implies that the right-hand side of (A.1) vanishes as $n \rightarrow \infty$, while the right-hand side of (A.2) vanishes as $n \rightarrow \infty$ thanks to Propositions 2. □

A.3. Lemma 1. In what follows, for each $n \geq 1$, let $\hat{\tau}_n$ be the context tree selected by applying Algorithm 2 to the sample data

$$(X_1, Y_1), \dots, (X_n, Y_n),$$

where for each $1 \leq m \leq n$, $Y_m = (Y_m^e : e \in \mathcal{E})$ for some finite set of indexes \mathcal{E} , and for $e \in \mathcal{E}$, the sample $(X_1, Y_1^e), \dots, (X_n, Y_n^e)$ is produced by a stochastic process driven by a context tree compatible with a probabilistic context tree $(\bar{\tau}, \bar{p})$ (recall the setup presented in Subsection (4.4)). The Theorem 1 reads the following.

LEMMA 1. *Given $\alpha \in (0, 1]$ and $n \geq 1$, there exist c_n , C_1 and C_2 such that*

$$P(\bar{\tau} \preceq \hat{\tau}_n) \leq \alpha.$$

PROOF. Given $\alpha_1 \in (0, 1]$, using Kolmogorov-Smirnov statistic properties we may find a threshold value c_n such that for any finite string $s \in A^*$ with $s \notin \bar{\tau}$ and any realization of a Brownian motion $W = (W(t) : t \in [0, T])$, it holds

$$P(\Delta_n^W(s) > c_n) \leq \alpha_1.$$

In the sequel we write $\tilde{\Delta}_n(s) = \tilde{\Delta}_n^e(s)$ to stress the dependence on $e \in \mathcal{E}$. By the independence of W_1, \dots, W_N , it follows that conditioning on Y_1^e, \dots, Y_n^e , the random variable $\tilde{\Delta}_n^e(s)$ is stochastically dominated by a Binomial random variable $B(N, \alpha_1)$. From this it is straightforward to deduce that for any given $\alpha_2 \in (0, 1]$, it is possible to find a constant C_1 such that for any finite string $s \notin \bar{\tau}$,

$$P(\tilde{\Delta}_n^e(s) > C_1) \leq \alpha_2.$$

Since for any $s \in A^*$, $\{s \in \hat{\tau}_n^e\} \subset \{\tilde{\Delta}_n^e(\text{suf}(s)) > C_1\}$ (recall the definition of $\hat{\tau}_n^e$ given in Subsection 4.4), it follows from the inequality above that for any $s \notin \bar{\tau}$,

$$P(s \in \hat{\tau}_n^e) \leq \alpha_2.$$

Notice that in Algorithm 2 all the Brownian random directions chosen to compute $\hat{\tau}_n^e$ and $\hat{\tau}_n^{e'}$ are independent for each $e, e' \in \mathcal{E}$. We may deduce (similarly as above) that conditioning on (Y_1, \dots, Y_n) , the random variable $Z_n(s)$ with $s \notin \bar{\tau}$ is stochastically bounded a Binomial random variable $B(|\mathcal{E}|, \alpha_2)$. It follows then that given any $\alpha \in (0, 1]$, we may chose C_2 in such a way that

$$P(\bar{\tau} \preceq \hat{\tau}_n) \leq \sum_{w \in \bar{\tau}} \sum_{\substack{s \succ w: \\ \ell(s) \leq L}} P(Z_n(s) > C_2) \leq \alpha.$$

□

ACKNOWLEDGEMENTS

We thank Marcelo Queiroz for synthesizing the acoustic stimuli. We also thank Sebastian Hofle for setting the experimental design and together with Erika Lazary and Douglas Rodrigues, collecting the experimental data. We also thank Manuel Febrero and Wenceslao Gonzalez Manteiga for a long discussion on functional data statistics.

This work is part of USP project *Mathematics, computation, language and the brain*, FAPESP project *Research, Innovation and Dissemination Center for Neuromathematics* (grant 2013/07699-0), projects *Stochastic modeling of the brain activity* (CNPq grant 480108/2012-9) and *Plasticity in the brain after a brachial plexus lesion* (CNPq grant 478537/2012-3 and FAPERJ grants E26/110.526/2012 and E26/010.002902/2014).

SUPPLEMENTARY MATERIAL

Supplement A: Data and Scripts

(<http://www.e-publications.org>). The directory SUPPLEMENT contains the subdirectories DATASET and SCRIPTS. The data files for each volunteer as well as a Readme file which contains the data description are available in the directory DATASET. The SCRIPTS directory contains three subdirectories PRE-PROCESSING, SIMULATION and PROJECTIVE-METHOD. In the PRE-PROCESSING directory is placed all the scripts concerning the pre-processing of the dataset, briefly described in Subsection 4.2. In the SIMULATION directory the scripts used in Section 5 are provided. The

PROJECTIVE-METHOD directory contains all the scripts used in the data analysis. Each one of these subdirectories include also a Readme files explaining how to use the scripts.

REFERENCES

- BÜHLMANN, P. and WYNER, A. J. (1999). Variable length Markov chains. *Ann. Statist.* **27** 480–513.
- CUESTA-ALBERTOS, J. A., FRAIMAN, R. and RANSFORD, T. (2006). Random projections and goodness-of-fit tests in infinite-dimensional spaces. *Bulletin of the Brazilian Mathematical Society, New Series* **37** 477–501.
- DELORME, A. and MAKEIG, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods* **134** 9–21.
- DOYA, K., ISHII, S., POUGET, A. and RAO, R. P. N. (2007). *Bayesian brain: Probabilistic approaches to neural coding*. The MIT Press.
- FERRATY, F. and ROMAIN, Y. (2011). *The Oxford Handbook of Functional Data Analysis*. Oxford University Press, USA.
- FRISTON, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **360** 815–836.
- FRISTON, K. (2012). The history of the future of the Bayesian brain. *NeuroImage* **62** 1230 – 1233.
- FRISTON, K. and KIEBEL, S. (2009). Predictive coding under the free-energy principle. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **364** 1211–1221.
- GALVES, A. and LÖCHERBACH, E. (2008). Stochastic chains with memory of variable length. *TICSP Series* **38** 117–133.
- GARRIDO, M. I., KILNER, J. M., KIEBEL, S. J. and FRISTON, K. J. (2009). Dynamic causal modeling of the response to frequency deviants. *J Neurophysiol* **101** 2620–2631.
- KERSTEN, D., MAMASSIAN, P. and YUILLE, A. L. (2004). Object perception as Bayesian inference. *Annual Rev. Psychology* **55** 271–304.
- KÖRDLING, K. P. and WOLPERT, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences* **10** 319–326.
- LEE, T. S. and MUMFORD, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America. A, Optics, image science, and vision* **20** 1434–1448.
- MEYER, T. and OLSON, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences* **108** 19401–19406.
- MUMFORD, D. (1992). On the Computational Architecture of the Neocortex, pt. II, The Role of the Cortico-Cortical Loop. *Biological Cybernetics* **65** 241–251.
- ORBÁN, G. and WOLPERT, D. M. (2011). Representations of uncertainty in sensorimotor control. *Current Opinion in Neurobiology* **21** 629 – 635. Sensory and motor systems.
- RISSANEN, J. (1983). A Universal Data Compression System. *IEEE Trans. Inf. Theor.* **29** 656–664.
- SUMMERFIELD, C., EGNER, T., GREENE, M., KOECHLIN, E., MANGELS, J. and HIRSCH, J. (2006). Predictive Codes for Forthcoming Perception in the Frontal Cortex. *Science* **314** 1311–1314.
- TENENBAUM, J. B., KEMP, C., GRIFFITHS, T. L. and GOODMAN, N. D. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. *Science* **331** 1279–1285.

- VON HELMHOLTZ, H. (1867). *Handbuch der physiologischen Optik* **III**. Leopold Voss translated by The Optical Society of America in 1924 from the third german edition, 1910, Treatise on physiological optics, Vol. III.
- WACONGNE, C., CHANGEUX, J. P. and DEHAENE, S. (2012). A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. *The Journal of Neuroscience* **32** 3665-3678.
- WACONGNE, C., LABYT, E., VAN WASSENHOVE, V., BEKINSCHTEIN, T., NACCACHE, L. and DEHAENE, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences* **108** 20754-20759.
- WOLPERT, D. M., DOYA, K. and KAWATO, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London* **358** 593-602.
- WOLPERT, D. M. and FLANAGAN, J. R. (2001). Motor prediction. *Current Biology* **11** R729-R732.
- WOLPERT, D. M., GHAHRAMANI, Z. and FLANAGAN, J. R. (2001). Perspectives and problems in motor learning. *Trends in Cognitive Sciences* **5** 487-494.

A. DUARTE

A. GALVES

G. OST

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
UNIVERSIDADE DE SÃO PAULO

BRAZIL

E-MAIL: aline.duart@gmail.com

E-MAIL: galves@usp.br

E-MAIL: guilhermeost@gmail.com

R. FRAIMAN

CENTRO DE MATEMÁTICA, FACULTAD DE CIENCIAS

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

URUGAY

E-MAIL: rfraiman@cmat.edu.uy

C. VARGAS

LABORATÓRIO DE NEUROBIOLOGIA II, INSTITUTO DE BIOFÍSICA CARLOS CHAGAS FILHO

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

BRAZIL

E-MAIL: cdvargas@biof.ufrj.br